

Sign Synthesis and Sign Phonology

Angus B. Grieve-Smith

Linguistics Department

University of New Mexico

Over the past half century, computerized speech synthesis has provided not only applications for linguistic theory, but also a source of feedback that allows those theories to grow. Recent developments in computer animation have now made sign synthesis possible with almost any personal computer manufactured today, and the same benefits will soon emerge for sign linguistics. In the process of developing SignSynth, an Internet-based prototype sign synthesis application, I have found it helpful to treat the phonology of American Sign Language (ASL) as being composed of at least four distinct subsystems.

The history of sign synthesis has paralleled the older field of speech synthesis. Just as the “pattern playback” machines of the 1950s produced understandable, if unnatural, speech from invented spectrograms painted on paper (Lieberman et al. 1957), the early sign synthesis programs produced understandable stick-figure movements based on detailed descriptions of the motions to be synthesized (Loomis, Poizner and Hollerbach 1982). Both of these systems involved a low-level correspondence between input and output. Current prototypes are more advanced, and reports of work in progress at the University of Delaware (Messing 1997) and Hitachi Central Research Laboratories (Ohki et al. 1994) indicate that at least those two take into account some aspects of sign phonology.

One major problem with many of the sign synthesis programs currently under development is their confusion of synthesis with translation. It is a well-established principle that a signed language is not simply a signed version of some spoken language. This principle is one of the foundations of sign linguistics, but it is often ignored in

discussions of sign synthesis. For example, the statement by one of the officers of the company Seamless Solutions that “We’re working on an animation engine that will take any text to sign language. [...] If a word isn’t in the current vocabulary, it will be communicated through finger spelling” (Wideman and Sims 1998) implies that English is the “text” form of American Sign Language. Even systems such as the Delaware and Hitachi prototypes, which clearly distinguish between ASL and English, and Japanese Sign Language (JSL) and Japanese, respectively, still focus primarily on word-for-word translation. In addition to perpetuating this myth of signed/spoken language equivalence, they also render the programs incapable of synthesizing signs which have no one-word equivalent in the spoken language.

One reason for the persistence of the myth of signed/spoken language equivalence is that the notion of text is not very well-developed for signed languages. In everyday written correspondence, most signers use some form of a spoken language. Many signed-language scholars use ad hoc “gloss” notations, whereby signs are represented in their natural order by uppercase words taken from a spoken language, along with invented forms such as “PRO.1” for some grammatical words. Other scholars use one of the nine phonologically-based notations (Miller 1994), such as Stokoe notation or HamNoSys.

For SignSynth, I felt it was important to use a phonologically-based notation as the input text, and output true ASL. As an initial starting point, I chose ASCII-Stokoe, Mark Mandel's (1993) adaptation of the notation used in the Dictionary of American Sign Language (Stokoe, Casterline and Croneberg 1965). This notation allows for the expression of almost all the lexical signs in ASL, but it is clear that there are still things left out.

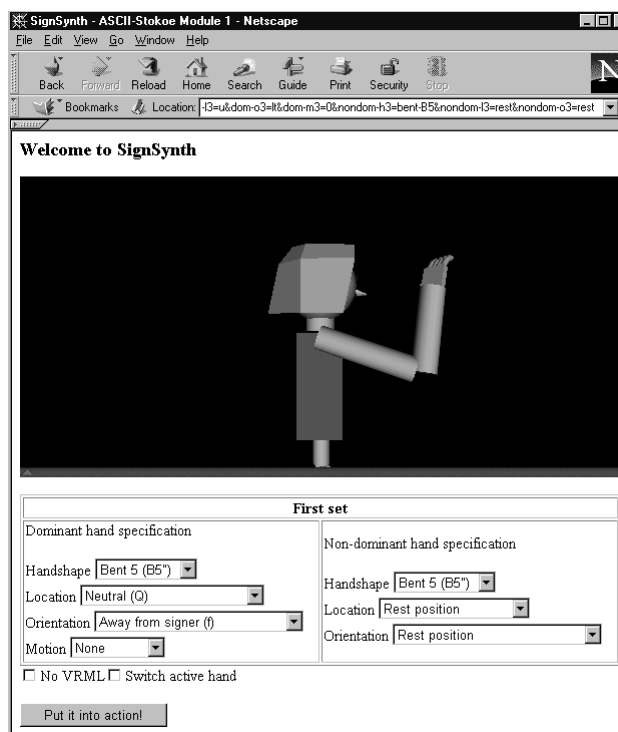


Figure 1. ASCII-Stokoe input module

In fact, in addition to the phonology that governs lexical signs there are at least three other phonological subsystems, all of which are used in a typical ASL conversation. The lexical signs, such as those for “man” or “eat,” were shown by Stokoe (1960) to draw on a finite set of options for the phonological parameters of handshape, location and movement. Fingerspelling is used mostly to produce words or names borrowed from spoken languages. Fingerspelled letters and numbers use a much smaller set of locations, but a larger set of handshapes. Classifier predicates use a very small set of handshapes, but have a wide range of locations and movements to describe in detail the size, shape and movement of objects. Finally, nonmanual gestures do not use handshapes, locations and movements, but are no less essential to the grammar of ASL. They are used to mark topic

and comment structure, questions and subordinate clauses. Some nonmanuals, such as the sign for “very interesting” stand for lexical items by themselves; others can distinguish one lexical item from another, for example the signs for “late” and “not yet” are identical except that the sign for “not yet” requires the signer’s tongue to be visible.

This division in the phonology of ASL is reflected in the interface to SignSynth. SignSynth has one module which allows the user to specify lexical signs in ASCII-Stokoe (see Figure 1). This module has a set of pull-down menus each containing the list of possibilities for a particular parameter. For example, the sign for “stuck” (ASCII-Stokoe k/Vt/x.) is specified by choosing the location (k for neck), handshape (V) and orientation (t for towards the signer) from each list. (Movement, the last parameter (x for contact, period for repetition), is currently handled by specifying a series of holds.) For a compound sign or a sequence of signs, the application can generate a form with a set of menus for each hold in the sequence.

The fingerspelling module provides a different interface to the ASCII-Stokoe module. It provides a field for the user to type text in the Roman alphabet, which is then converted into an animation sequence. The user can also control the speed and handedness of the fingerspelling. The module is able to synthesize the letters J and Z, which are formed using more than one hold, but does not produce coarticulated fingerspelling or loan signs. It also does not reproduce the common practice of representing double letters by extending the hold and moving the hand outward.

The other two subsystems have not been fully implemented yet, and will likely be integrated with the first two. SignSynth can already produce a few nonmanual gestures: in the ASCII-Stokoe module it is possible to specify eyegaze and eyebrow position to be

articulated simultaneously with a particular sign. Classifier predicates are not yet supported.

Once this information is specified, the application converts the phonological specification into a set of keyframe interpolators in Virtual Reality Modeling Language (VRML). These interpolators can be used, with slight modification, to animate any virtual humanoid that has arms and fingers and is compliant with the standards of the VRML Humanoid Animation Working Group. SignSynth currently includes two compliant humanoids, and the application will attach the interpolators to one of the humanoids to produce a self-contained VRML file, which can be displayed on most of the popular operating systems currently in use. This usually happens automatically: after submitting the phonological specification, the initial screen of the VRML file is displayed. The user can then click the mouse button to display the animation as many times as desired, or save the file for later viewing. VRML also allows the user to rotate the virtual signer and view the sign from any angle. Finally, the file can be converted to a variety of digitized video and image formats and displayed, emailed or printed.

There are several potential applications for this technology. First, synthesized speech has been used as a stimulus in several important psycholinguistic experiments, such as categorical perception and gating. Synthesized sign would be able to provide similar controlled stimuli to see if the findings of these psycholinguistic experiments pertained specifically to speech, or rather to language in general. Second, many online dictionaries of signed languages use large collections of digitized video clips for examples; these dictionaries would be easier to produce and store if they only had to specify a short ASCII-Stokoe or HamNoSys representation; the articulation of the example could be

done using a system such as SignSynth. Also, synthesized sign could be used in pedagogical applications such as flash cards and quiz games.

As it is now, SignSynth is far from complete, and there are several directions in which it could be extended. First, there are problems relating to the modeling of anatomical motion which prevent SignSynth from realizing the full range of lexical signs; this is an important obstacle to overcome. I am also currently working to provide support for classifier predicates and nonmanual gestures. Another enhancement will be a parser to allow users familiar with ASCII-Stokoe notation to type a sequence of signs in directly, rather than choosing specifications from menus; this would also flag fingerspelled words and route them to the appropriate module. Later, I would like to add support for other notation systems, such as SignFont and HamNoSys.

In this discussion I have demonstrated SignSynth, a prototype sign synthesis program, and discussed some of the linguistic principles it has brought to light. The prototype of SignSynth is currently available as a free application on the UNM Linguistics Department's World Wide Web server at <<http://s-leodm.unm.edu/signsynth>>. Everyone is invited to try it, and comments and questions are welcome.

Bibliography

- LIBERMAN, ALVIN M., KATHERINE SAFFORD HARRIS, HOWARD S. HOFFMAN AND BELVER C. GRIFFITH. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54, 358-368.
- LOOMIS, JEFFREY, HOWARD POIZNER AND JOHN HOLLERBACH. 1983. Computer graphic modeling of American Sign Language. *Computer Graphics* 17, 105-114.
- MANDEL, MARK A. 1993. ASCII-Stokoe notation: A computer-writeable transliteration system for Stokoe notation of American Sign Language. Unpublished manuscript.
- MESSING, LYNN AND GARLAND STERN. 1997. Sister Mary article. Unpublished manuscript.
- MILLER, CHRISTOPHER. 1984. A note on notation. *Signpost* 7, .

- OHKI, MASARU, HIROHIKO SAGAWA, TOMOKO SAKIYAMA, EIJI OOHIRA, HISAH I KEDA AND HIROMICHI FUJISAWA. 1994. Pattern recognition and synthesis for sign language translation system. ASSETS 10, 1-8.
- STOKOE, WILLIAM C. 1960. Sign language structure: An outline of the visual communication systems of the American Deaf. Studies in Linguistics, Occasional Papers 8.
- STOKOE, WILLIAM C., DOROTHY C. CASTERLINE AND CARL G. CRONEBERG. 1965. A dictionary of American Sign Language on linguistic principles. Silver Spring: Linstok Press.
- WIDEMAN, CAROL. 1998. Sign language. h-anim@vrml.org (April 22, 1998). <<http://www.vrml.org/WorkingGroups/h-anim/hypermail/1998/0193.html>>